CPFNet: Context Pyramid Fusion Network for Medical Image Segmentation

Shuanglang Feng, Heming Zhao, Fei Shi, Xuena Cheng, Meng Wang, Yuhui Ma, Dehui Xiang, Weifang Zhu, Xinjian Chen, *Senior Member, IEEE*

***Abstract*—Accurate and automatic segmentation of medical images is a crucial step for clinical diagnosis and analysis. The convolutional neural network (CNN) approaches based on the U-shape structure have achieved remarkable performances in many different medical image segmentation tasks. However, the context information extraction capability of single stage is insufficient in this structure, due to the problems such as imbalanced class and blurred boundary. In this paper, we propose a novel Context Pyramid Fusion Network (named CPFNet) by combining two pyramidal modules to fuse global/multi-scale context information. Based on the U-shape structure, we first design multiple global pyramid guidance (GPG) modules between the encoder and the decoder, aiming at providing different levels of global context information for the decoder by reconstructing skip-connection. We further design a scale-aware pyramid fusion (SAPF) module to dynamically fuse multi-scale context information in high-level features. These two pyramidal modules can exploit and fuse rich context information progressively. Experimental results show that our proposed method is very competitive with other state-of-the-art methods on four different challenging tasks, including skin lesion segmentation, retinal linear lesion segmentation, multi-class segmentation of thoracic organs at risk and multi-class segmentation of retinal edema lesions.**

***Index Terms*—Medical image segmentation, convolutional neural network, context pyramid fusion network, global pyramid guidance module, scale-aware pyramid fusion module**

I. Introduction

THE semantic segmentation of medical images plays an important role in medical image analysis, such as skin lesion segmentation in dermoscopy images [1], [2], retinal linear lesion segmentation in indocyanine green angiography (ICGA) images [3], segmentation of thoracic organs at risk in computed tomography (CT) images [4], and retinal edema lesion segmentation in optical coherence tomography (OCT) images [5], [6]. Accurate and automatic target segmentation can be used to derive quantitative assessment of pathology or biomarkers for subsequent diagnosis, treatment planning and disease progression monitoring.

Recently, many deep learning methods based on convolutional neural networks (CNN) have been applied to medical image segmentation tasks because of their excellent capability of feature extraction [7], [8], [9], [10].

In CNN framework consisting of stacked convolutional layers and downsampling layers, deeper stages usually have wider range of receptive fields and are able to capture global context information, while shallower stages usually have local information with higher spatial resolution features. Based on these, many new structures based on fully convolutional network (FCN) were proposed for semantic segmentation tasks [7], [11], [12], [13]. Among them, U-Net [7] has achieved remarkable performances. In the encoder-decoder structure represented by U-Net, an original FCN is employed as the encoder to capture high-level semantic information gradually by stacking convolutional layers and downsampling layers. A down-top decoder is designed to recover the spatial information from the output of encoder stage by stage. Meanwhile, multiple skip-connections between decoder and encoder are utilized to make up for the fine information loss caused by downsampling, which improve the performance significantly.

Although CNNs with U-shape structures have achieved remarkable performances and received a lot of attention in many medical image segmentation applications [10], [14], [15], [16] for each single stage, the capability of context information extraction is still insufficient.

First, on one hand, the global context information captured by deeper stages of the encoder is gradually transmitted to shallower layers, which may be progressively diluted since the feature extraction ability of a single stage is weak. On the other hand, the simple skip-connection in each stage ignores global information and is an indiscriminate combination of local information that will introduce irrelevant clutters and result in misclassification of pixels. Recently, some approaches have been proposed to try to solve these problems. FastFCN [17] used a joint pyramid upsampling module to replace dilated

Manuscript received August 11, 2019. This study was supported in part by the National Key R&D Program of China under Grant 2018YFA0701700, in part by the National Nature Science Foundation of China (61622114, 81401472) and in part by the National Basic Research Program of China (2014CB748600) and the International Cooperation Project of Ministry of Science and Technology (2016YFE010770). Shuanglang Feng and Heming Zhao contributed equally to this work. Corresponding authors: Weifang Zhu (wfzhu@suda.edu.cn) and Xinjian Chen (xjchen@suda.edu.cn).

Shuanglang Feng, Heming Zhao, Fei Shi, Xuena Cheng, Wang Meng, Yuhui Ma, Dehui Xiang and Weifang Zhu are with the School of Electronics and Information Engineering, Soochow University, Jiangsu 215006, China (email: 2470697802@qq.com, hmzhao@suda.edu.cn, shifei@suda.edu.cn, 834614265@qq.com, wangmeng9218@126.com, 815349387@qq.com, xiangdehui@suda.edu.cn and wfzhu@suda.edu.cn).

Xinjian Chen is with the School of Electronics and Information Engineering and the State Key Laboratory of Radiation Medicine and Protection, Soochow University, Jiangsu 215006, China (email: xjchen@suda.edu.cn).

convolutions and capture global context information. Anatomynet [18] and DFN [19] adopted channel attention mechanism to guide shallow stages to learn global feature representations. GCN [20] and MultiResUNet [2] added a larger kernel and deeper convolution layer respectively in skip-connection to transform local semantic information to higher-level features. Attention U-Net [21] utilized a novel attention gate (AG) module to highlight salient features that are useful for a particular task. However, there are few methods to solve both problems simultaneously.

Second, in each single stage, there is no effective extraction and utilization of multi-scale context information. When dealing with targets with complex structures, such information is necessary so that the structure's surroundings are also considered and ambiguous decisions can be avoided [22]. Recently, some methods have been proposed to explore and integrate multi-scale context information. PSPNet [23] and PoolNet [24] adopted multiple parallel poolings with different kernel sizes to process high-level feature maps. DeepLab v3 [25] and CE-Net [26] adopted multiple convolution branches with different receptive fields to improve the multi-scale information capture ability of the model. However, in their methods, the receptive fields cannot be dynamically adjusted to fit the targets with different sizes. Since attention mechanism [19],[27] has been widely used for improving model performance, many scale-aware networks based on attention mechanism have been proposed to overcome the above problems. SA[28] learned to softly weight the multi-scale features of each pixel by introducing attention module to multi-scale inputs. AFNet [22] and SPAP [29] employed scale-aware layers to adaptively change the sizes of the effective receptive fields. SKNet [30] proposed a dynamic kernel selection mechanism by employing channel attention mechanism into multiple feature branches.

In this paper, we introduce two novel pyramidal modules into U-shape network to solve the aforementioned problems. Motivated by the discussion of the first problem, we design a Global Pyramid Guidance (GPG) module, which combines multi-stage global context information to reconstruct skip-connection and provide global information guidance flow for the decoder. Specifically, each stage ' s skip-connection consists of both local context information from this stage and global context information from higher-level stages. Meanwhile, by introducing GPG, irrelevant background noise brought by low-level features can be suppressed. Motivated by the discussion of the second problem and the scale-aware mechanism, we further propose a Scale-Aware Pyramid Fusion (SAPF) module, which consists of three parallel dilated convolutional filters with shared weights for capturing different scale context information and two cascaded Scale-Aware Modules (SAMs) employing the spatial attention mechanism. The SAPF module is embedded at the top of the backbone, which can dynamically select the appropriate receptive fields for targets with different scales by self-learning and fuse multi-scale context information better.

Based on the above descriptions, we term our method Context Pyramid Fusion Network (CPFNet). The proposed CPFNet is applied for four challenging medical image segmentation tasks: skin lesion segmentation in dermoscopy images, retinal linear lesion segmentation in ICGA images, thoracic risky organs segmentation in CT images and retinal edema lesions segmentation in OCT images. The second task is based on clinical datasets while the first, third and fourth tasks are based on public benchmark datasets.

Our main contributions are summarized in three aspects as follows:

(1) Two novel pyramidal modules including GPG module and SAPF module are proposed to effectively fuse global and multi-scale context information, respectively.

(2) Based on a U-shape network, the proposed GPG module and SAPF module can be easily embedded and applied for medical image segmentation tasks.

(3) The state-of-the-art segmentation performances for four different challenging tasks show that the proposed CPFNet has good generalization ability.

II. METHODS

A. Overview

Fig.1 demonstrates the proposed CPFNet, which is an FCN based on encoder-decoder architecture and consists of four main parts: feature encoder, GPG module, SAPF module and feature decoder. The SAPF module is inserted at the top of the encoder to capture multi-scale context information, while multiple GPG modules are placed between the encoder and the decoder to guide the fusion of the global context information flows and decoder path features.

B. Feature Encoder

In order to get more representative feature maps, we employ a pre-trained ResNet34 [31] as the feature extractor. For compatibility purpose, the average pooling layer and fully connected layers are removed. Because of the residual blocks with shortcut mechanism, as shown in bottom right of Fig.1, the ResNet can accelerate convergence of the network and avoid gradient vanishing.

C. Global Pyramid Guidance module

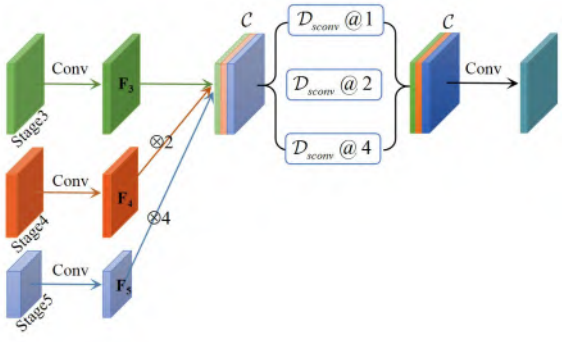

Fig.2. Illustration of the Global Pyramid Guidance (GPG) module. Taking the reconstructed skip-connection on Stage3 as an example, the global information flow is transmitted to the decoder by fusing the global context information from higher stages (Stage4 and Stage5).

From the input image, the encoder can learn global context information including the object's surroundings and the

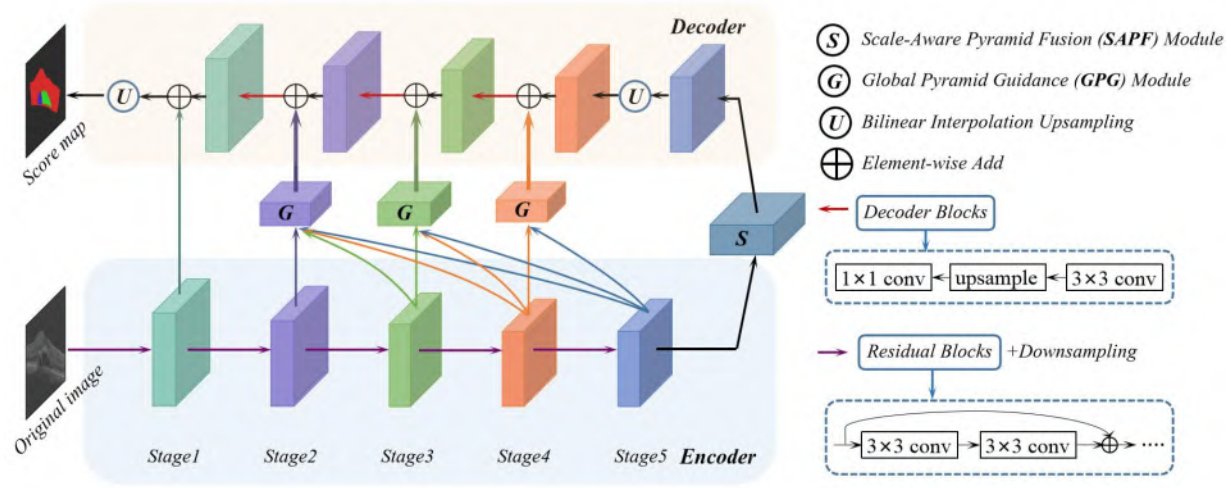

Fig.1. Overview of the proposed CPFNet. The original image is fed into the encoder composed of pre-trained ResNet34 to obtain the high-level features, and then the multi-scale information is captured and dynamically merged by the proposed scale-aware pyramid fusion (SAPF) module. Next, the features are recovered by the decoder and meanwhile the global context information flows are introduced by the proposed global pyramid guidance (GPG) module. Here, the decoder consists of 3×3 convolution, bilinear interpolation up-sampling, and 1×1 convolution. Finally, the predicted score map is obtained.

category characteristics of the object [26], [32]. However, these types of information may be progressively weakened when they are gradually transmitted to shallower layers [24]. Besides, the original skip-connection in the U-shape network will introduce irrelevant clutters and have semantic gap due to the mismatch of receptive fields. In this paper, we propose a global pyramid guidance (GPG) module to solve these problems, which is shown in Fig.2.

In the GPG module, the skip-connection is reconstructed by combining the feature map of this stage with the feature maps of all higher-level stages. For example, Fig.2 shows the GPG module on Stage 3. First, features of all stages are mapped into the same channel space as Stage3 by a regular 3×3 convolution. Next, the generated feature maps $\mathbf{F}_4$ and $\mathbf{F}_5$ are upsampled to the same size as $\mathbf{F}_3$ and concatenated. Then in order to extract global context information from different levels of feature maps, three separable convolutions [33]($\mathcal{D}_{sconv}@1$, $\mathcal{D}_{sconv}@2$, $\mathcal{D}_{sconv}@4$)with different dilation rates (1, 2 and 4) are employed in parallel, where separable convolutions are used to reduce model parameters. It's worth noting that the number of parallel paths and dilation rates vary with the number of fused stages. Finally, a regular convolution is employed to obtain the final feature map. Above all, each GPG module in different stages can be summarized as (to simplify the formula, regular convolution is ignored):

$$\mathbf{G}_k = \underset{i=k}{\overset{i=5}{\mathcal{C}}}(\mathcal{D}_{sconv}@2^{i-k}(\underset{i=k}{\overset{i=5}{\mathcal{C}}}(\mathbf{F}_k \otimes 2^{i-k}))) \quad (1)$$

Where $\mathbf{G}_k$ denotes the output of GPG module inserted in the k^{th} stage, $\mathbf{F}_k$ denotes the feature map of the k^{th} stage in the encoder, $\otimes 2^{i-k}$ represents the upsampling operation with rate of 2^{i-k}, $\mathcal{C}$ represents the operation of concatenation and $\mathcal{D}_{sconv}@2^{i-k}$ represents the separable dilated convolution with dilation rate of 2^{i-k}.

To reduce the cost of computation, only three GPG modules are used in our network. By introducing multiple GPG modules between encoder and decoder, the global semantic information flow from high-level stages can be gradually guided to different stages.

D. Scale-Aware Pyramid Fusion module

As has been discussed in the introduction, multi-scale context information can improve the performance of semantic segmentation tasks. However, how to effectively integrate such information is a problem worth exploring. Inspired by this problem, we propose a Scale-Aware Pyramid Fusion (SAPF) module which is shown in Fig.3. In the SAPF module, we use three parallel dilated convolutions with different dilation rates of 1, 2 and 4 to capture different scale information. Note that these different dilated convolutions have shared weights, which can reduce the number of model parameters and the risk of overfitting.

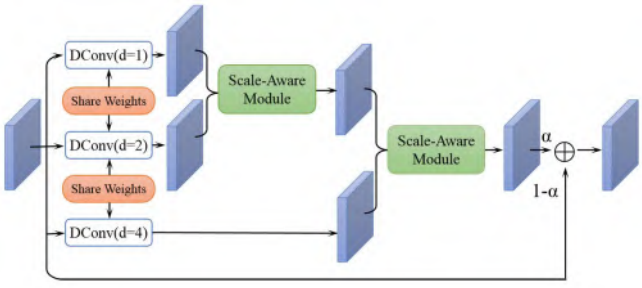

Fig.3. The illustration of Scale-Aware Pyramid Fusion (SAPF) module. Multi-scale information captured by three parallel dilated convolution layers with shared weights is dynamically fused by two scale-aware modules.

After this, we design a scale-aware module to fuse different scale features. As shown in Fig.4, a spatial attention mechanism is introduced to dynamically select the appropriate scale features and fuse them by self-learning. Specifically, two different scale features $\mathbf{F}_A$ and $\mathbf{F}_B$ pass through a series of

convolutions and obtain two feature maps $\mathbf{A}$, $\mathbf{B}$ $\in \mathbb{R}^{H\times W}$ (H: the height of feature map, W: the width of feature map). Then pixel-wise attention maps $\mathcal{A}$, $\mathcal{B}$ $\in \mathbb{R}^{H\times W}$ are generated by softmax operator on the spatial-wise values:

$$\mathcal{A}_i = \frac{e^{\mathbf{A}_i}}{e^{\mathbf{B}_i}+e^{\mathbf{A}_i}}, \mathcal{B}_i = \frac{e^{\mathbf{B}_i}}{e^{\mathbf{B}_i}+e^{\mathbf{A}_i}}, \quad i=[1,2,3...,H\times W] \tag{2}$$

Finally, the fusion feature map is obtained as a weighted sum:

$$\mathbf{F}_{fusion} = \mathcal{A}\odot\mathbf{F}_A + \mathcal{B}\odot\mathbf{F}_B \tag{3}$$

where the element-wise product operations ($\odot$) are performed between the attention maps and two scale features to get the fused feature map $\mathbf{F}_{fusion}$.

We employ two cascaded scale-aware modules to get the final fusion feature of three branches. Then a residual connection with learnable parameter α is employed to obtain the output of the whole SAPF module.

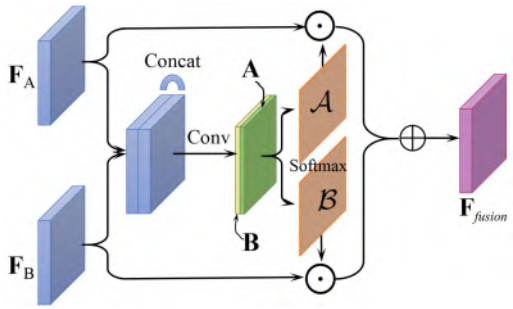

Fig.4. The illustration of scale-aware module. Different weight maps are applied to the features of different receptive fields through the spatial attention mechanism.

E. Feature Decoder

In order to restore high resolution feature maps quickly and efficiently, multiple simple decoder blocks are used in the decoder path. The decoder restores the spatial information with high-level features generated by the SAPF module, and gradually fuses the global context information guided by the GPG module via a 3×3 convolution as shown in Fig.1. Following the 3×3 convolution, a bilinear interpolation is used to upsample the fused feature maps, which can reduce the parameters of the model and checkerboard artifacts [34]. The output of a decoder block is obtained after 1×1 convolution. Note that after the last decoder block, the feature map is directly upsampled to the same size with the original input image.

F. Loss Function

A main challenge in medical image segmentation is class distribution imbalance. In order to optimize our model further, we employ a joint loss $\mathcal{L}_{total}$ consisting of Dice loss $\mathcal{L}_{Dice}$ and cross-entropy loss $\mathcal{L}_{CE}$ to perform all segmentation tasks. The formula is as follows,

$$\mathcal{L}_{total} = \mathcal{L}_{Dice} + \lambda\mathcal{L}_{CE} \tag{4}$$

where λ is a trade-off between Dice loss and cross-entropy loss , and is set to 1 in all our experiments. For fair comparison, all methods in our experiments use the same loss function in each individual task.

G. Implementation Details

The encoder of our proposed model is based on the pre-trained ResNet34. The implementation of the proposed CPFNet is based on the public platform PyTorch and NVIDIA Tesla K40 GPU with 12GB memory. We use the 'poly' learning rate policy, where $lr = baselr\times(1-\frac{iter}{total\_iter})^{power}$, the basic learning rate $baselr$ is set to 0.01, and $power$ is set to 0.9. Batch size and iteration number vary according to the datasets. Besides, stochastic gradient descent (SGD) is adopted to optimize our model, in which momentum and weight decay are set to 0.9 and 0.0001 respectively. We will release our codes on Github[1].

III. EXPERIMENTS AND RESULTS

A. Skin Lesion Segmentation

1) Overview

Dermoscopy is a non-invasive imaging method which is widely used in clinical dermatology [35]. Skin lesion segmentation in dermoscopy images is of great value for automatic screening and detection of melanoma. Some traditional approaches have been proposed to analyze the dermoscopy images and segment the melanomas, including clustering, thresholding and region-based active contour models [36]. Sarker *et al.* [1] proposed an end point error loss and negative log-likelihood loss based on CNN to perform the skin lesion segmentation. MultiResUNet [2] introduced multiple residual connection in U-Net for dermoscopy image segmentation. However, there are still many challenges due to the inhomogeneity of dermoscopy images, the influence of dense hair and the blurred boundaries of lesions.

2) Dataset

The demoscopy image dataset was acquired from a public challenge: Lesion Boundary Segmentation in ISIC-2018[2]. The data for this challenge were extracted from the ISIC-2017 dataset [37] and the HAM10000 dataset [38], which were collected from different leading clinical centers internationally and acquired from different types of devices. The dataset includes 2594 images including different types of skin lesions with different resolutions. To improve the computational efficiency of the model, we resized the image to 256×192 while maintaining the average aspect ratio. Online random left-right flipping was applied for data augmentation.

3) Evaluation metrics

We performed 5-fold cross validation both in ablation experiments and contrast experiments. To evaluate the performance of model objectively, three official evaluation metrics in the challenge, including Jaccard index (Jac), Dice coefficient (Dice) and Accuracy (Acc) were adopted.

[1]https://github.com/FENGShuanglang/CPFNet_Project
[2]https://challenge2018.isic-archive.com/task1

TABLE I
THE RESULT OF CONTRAST EXPERIMENTS AND ABLATION STUDIES ON SKIN LESION SEGMENTATION TASK (MEAN ± STANDARD DEVIATION)

Methods	Jaccard(%)	Dice(%)	Accuracy(%)
FCN [11]	76.62±0.315	84.92±0.425	94.52±0.303
DFN [19]	77.83±0.602	86.01±0.579	94.92±0.531
UNet[7]	78.70±0.317	86.58±0.376	95.06±0.251
CE-Net [26]	79.99±0.855	87.50±0.733	95.46±0.616
Attention U-Net [21]	80.26±0.442	87.06±0.501	95.33±0.325
MultiResUNet [2]	80.30±0.372	\	\
FastFCN [17]	81.71±0.740	88.98±0.690	96.71±0.610
GCN [20]	82.15±0.810	89.19±0.613	97.13±0.572
Baseline	81.12±0.551	87.90±0.561	95.69±0.617
Baseline+SAPF_w/o_Dc	81.56±0.353	88.25±0.433	95.80±0.305
Baseline+SAPF_w/o_SA	81.79±0.425	88.46±0.532	95.92±0.359
Baseline+SAPF	82.15±0.328	88.88±0.390	96.00±0.228
Baseline-Wide	81.73±0.376	88.41±0.438	95.90±0.342
Baseline+GPG	82.26±0.415	89.16±0.449	96.02±0.191
Baseline+GPG_w/o_D_S	81.85±0.402	88.69±0.472	95.92±0.379
CPFNet	**82.86±0.421**	**89.89±0.510**	**96.30±0.206**

4) Results

As show in Table I, we compare our method with other excellent CNN based methods, including FCN [11], U-Net [7], Attention U-Net [21], FastFCN [17], CE-Net [26], GCN [20], DFN [19] and MultiResUNet (result as reported in [2]). Besides, in order to verify the validity of the proposed GPG module and SAPF module, we also conduct a series of ablation experiments. For convenience, we call the basic U-shape model with pre-trained ResNet34 backbone as the Baseline method.

Compared to FCN, U-Net achieves an increase of more than 2% for the main evaluation metric Jaccard index, which benefits from skip-connections. Similarly, MultiResUNet achieves a further improvement by matching the receptive fields of encoder and decoder features on skip-connections. It is worth noting that the proposed CPFNet achieves better performance than all of the above methods. Compared with the Baseline, the performance of the proposed CPFNet gets an overall improvement (1.74% for Jaccard index, 1.99% for Dice coefficient and 0.61% for Accuracy). The performance of GCN is comparable with the proposed CPFNet for Jaccard index, while CE-Net performs bad in this segmentation task. Fig.7 shows the visualization results of different models.

Ablation study for GPG: As shown in Table I, the addition of the proposed GPG modules (Baseline+GPG) achieves substantial improvement over the Baseline in terms of all three evaluation metrics. Meanwhile, the performance of GPG modules without separable dilated convolutions (GPG_w/o_D_S) is worse than complete GPG, which proves that parallel branches with different receptive fields are more conducive for global information acquisition. To further validate the effectiveness of the GPG modules, we compare the output of simple skip-connections and our GPG modules by means of feature map visualization. As can be seen from the Fig.5, compared with simple skip-connection, the global context information flow from the GPG module results in better response of the segmentation target, which greatly improves the segmentation accuracy.

Ablation study for SAPF module: As shown in Table II, the embedding of SAPF module into Baseline (Baseline+SAPF) also helps to improve the performance. Compared with the Baseline, the Jaccard index increases 1.03% and reaches 82.15%. Meanwhile, the Dice and Accuracy increase from 87.90%, 95.69% to 88.88%, 96.00% respectively, which benefits from the fact that the proposed SAPF module can dynamically fuse multi-scale context information. To further verify this point, we first insert a SAPF module without dilation convolution (SAPF_w/o_Dc) in the Baseline, and the Jaccard index decreases 0.59% than the complete SAPF module, which implies that the capture of multi-scale context information is necessary. Second, we insert a SAPF module without Scale-Aware module (SAPF_w/o_SA) in the Baseline, which also results in a performance reduction of 0.36% compared with the complete SAPF module and indicates that dynamical selection of multi-scale contextual information is more conducive for lesion segmentation. All of these results prove that our SAPF module can improve the segmentation performance of the network by combining both advantages of scale-aware mechanism and multi-scale context information fusion.

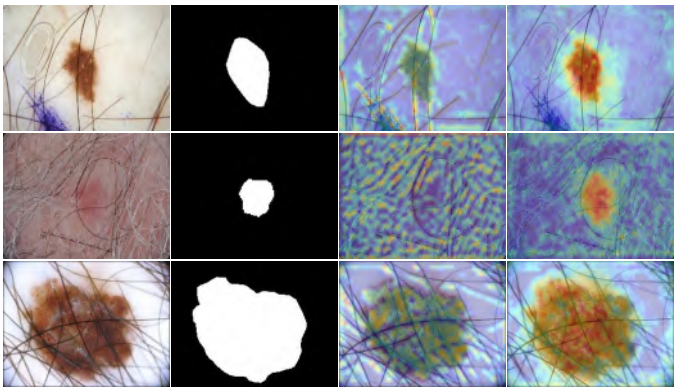

Fig.5. Comparison of feature maps transferred by the skip connection before and after insertion of GPG module. (a) Original image, (b) ground truth, (c) the feature map before inserting GPG modules (common skip-connection) and (d) the feature map after inserting GPG modules

Ablation study for model complexity and pre-training model: To verify that the performance improvement of our proposed model is not caused by increasing the model complexity, we design a network based on the Baseline with similar complexity to the CPFNet by adding multiple residual blocks in decoder, which is called Baseline-Wide in Table I. The experiments show that our proposed CPFNet achieves notable improvement than the Baseline-Wide (1.13% in term of Jaccard index). Besides, our Baseline also performs better than

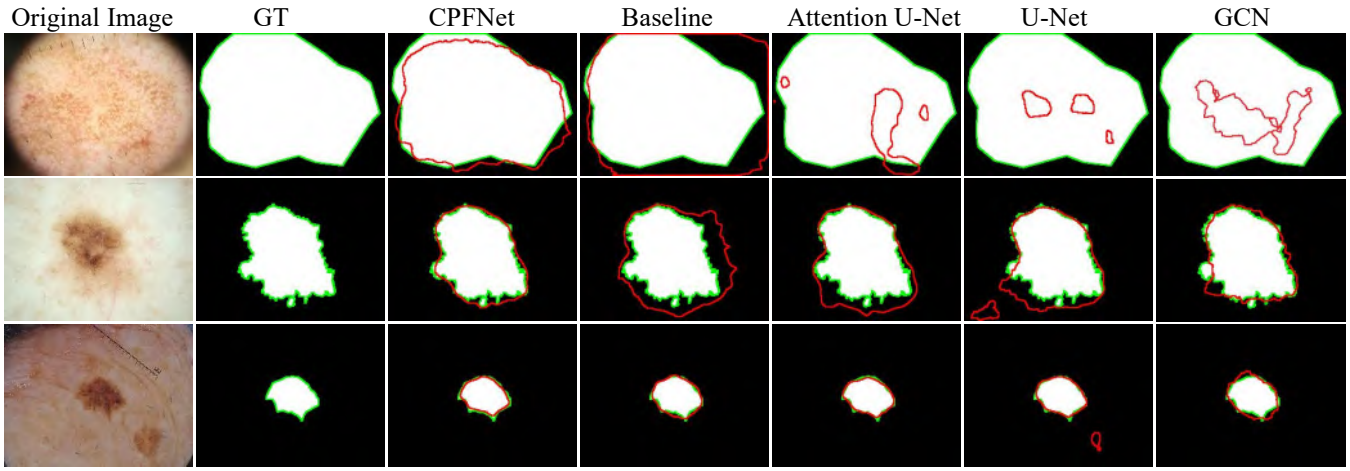

Fig.7. The examples of skin lesion segmentation. The areas outlined with green and red lines represent the ground truth and the prediction results, respectively. From left to right: original image, ground truth (GT), our CPFNet, Baseline, U-Net and GCN.

other methods listed in Table I, which benefits from the fact that ResNet34 with pre-trained weights makes the network easier to optimize and converge faster than that from scratch (shown in Fig.6) and the model is more powerful to capture useful features.

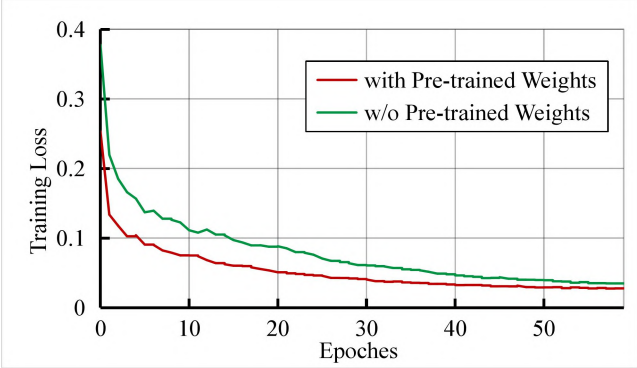

Fig.6. Comparison between the network with pre-trained weights and the one trained from scratch.

B. *Retinal Linear Lesion Segmentation*

1) *Overview*

Retinal linear lesions, including lacquer cracks and myopic stretch lines, are important indicators for the progress of high myopia [39], [40], which is a major cause of visual impairment. Indocyanine green angiography (ICGA) has been widely used for linear lesion examination in clinical ophthalmology. The linear lesion segmentation in ICGA image is crucial for the prevention and diagnosis of high myopia. However, it is very challenging due to the complex structures of the lesions and the similar characteristics with the retinal vessels. To the best of our knowledge, there are few studies focusing on automatic linear lesion segmentation. Jiang *et al.* [3] proposed an improved conditional generative adversarial (cGAN) network for linear lesion segmentation and achieved good performance. But there still are some drawbacks such as the high complexity of the improved cGAN model and too many hyperparamters.

2) *Dataset*

ICGA images (indocyanine green as fluorescer, Heidelberg Retina Angiography 2, Heidelberg Engineering, Heidelberg, Germany, 768 × 768 pixels) from 38 subjects (totally 76 eyes) with linear lesions were included, which were collected in Shanghai General Hospital from April 2017 to August 2017. The collection and analysis of images were approved by the Institutional Review Board of Shanghai General Hospital and adhered to the tenets of the Declaration of Helsinki. Because the number of subjects was small, two images collected at different time from each eye were included in the dataset. Multiple online random augmentation methods were used for data augmentation, including left-right flipping, up-down flipping, rotations from -30 degrees to 30 degrees and additive Gaussian noise additions.

3) *Evaluation metrics*

We divided the data into 4 folds according to subjects and conducted cross-validation. Jaccard index (Jac), Dice coefficient (Dice), Accuracy (Acc), Sensitivity (Sen) and Specificity (Spe) are adopted to verify the reliability of our method.

4) *Results*

First, in order to verify the versatility of the proposed GPG and SAPF modules, we insert these two modules into the original U-Net [7]. As shown in Table II, with the successive insertions of these two modules, the corresponding performances are stably improved. When both of the two modules are inserted, the Dice coefficient increases by nearly 9% and reaches 58.47%, and the Sensitivity increases by 11.57%. A U-Net-Wide network is designed by increasing the number of channels, which has the similar parameter number as U-Net+GPG+SAPF. The corresponding results in Table II show that the remarkable improvement of segmentation performance is not caused by the increase of parameters and strongly demonstrate that our proposed modules can make up for the weakness of context information capture ability in U-Net.

Second, we compare the proposed CPFNet with the state-of-the-art algorithms, including PSPNet [23], DFN [19]

TABLE II
THE RESULTS OF CONTRAST EXPERIMENTS AND ABLATION STUDIES ON RETINAL LINEAR LESION SEGMENTATION TASK (MEAN ± STANDARD DEVIATION)

Methods	IoU(%)	Dice(%)	Accuracy(%)	Sensitivity(%)	Specificity(%)
U-Net [7]	36.35±2.08	49.58±3.02	98.71±0.05	50.84±2.86	99.39±0.14
U-Net+GPG	40.66±2.88	55.98±3.47	98.87±0.12	56.73±5.33	99.41±0.09
U-Net+SAPF	41.69±2.13	56.52±2.46	98.83±0.09	60.49±5.36	99.44±0.15
U-Net+GPG+SAPF	43.14±3.35	58.47±4.25	98.89±0.27	62.41±6.20	99.22±0.19
U-Net-Wide	38.86±2.68	53.94±3.04	98.81±0.10	54.63±5.65	99.44±0.12
FastFCN [17]	25.62±2.53	39.03±3.37	98.62±0.10	33.23±3.56	99.68±0.06
PSPNet [23]	35.98±5.44	52.69±5.86	97.80±0.27	51.42±9.68	98.96±0.20
DFN [19]	37.14±4.01	54.18±4.72	98.83±0.18	54.18±4.21	**99.79±0.18**
MultiResUNet [2]	37.79±3.03	51.39±3.37	98.85±0.12	51.81±2.38	99.52±0.19
Attention U-Net [21]	40.11±3.08	54.86±4.15	98.87±0.07	54.97±5.09	99.48±0.12
TiramisuNet [41]	42.15±4.92	59.14±4.84	98.08±0.34	57.44±6.89	99.09±0.25
cGAN [42]	42.61±5.13	59.24±5.19	98.58±0.32	67.16±8.76	99.16±0.26
GCN [20]	43.07±3.73	59.13±3.61	98.92±0.08	56.53±5.31	99.53±0.05
CE-Net [26]	47.21±4.09	62.98±3.95	99.00±0.08	61.59±6.37	99.55±0.13
Baseline	44.25±5.07	59.37±2.63	98.91±0.13	68.53±7.69	99.38±0.17
CPFNet	**47.75±3.10**	**63.08±3.55**	**99.08±0.05**	**71.77±6.39**	99.67±0.04

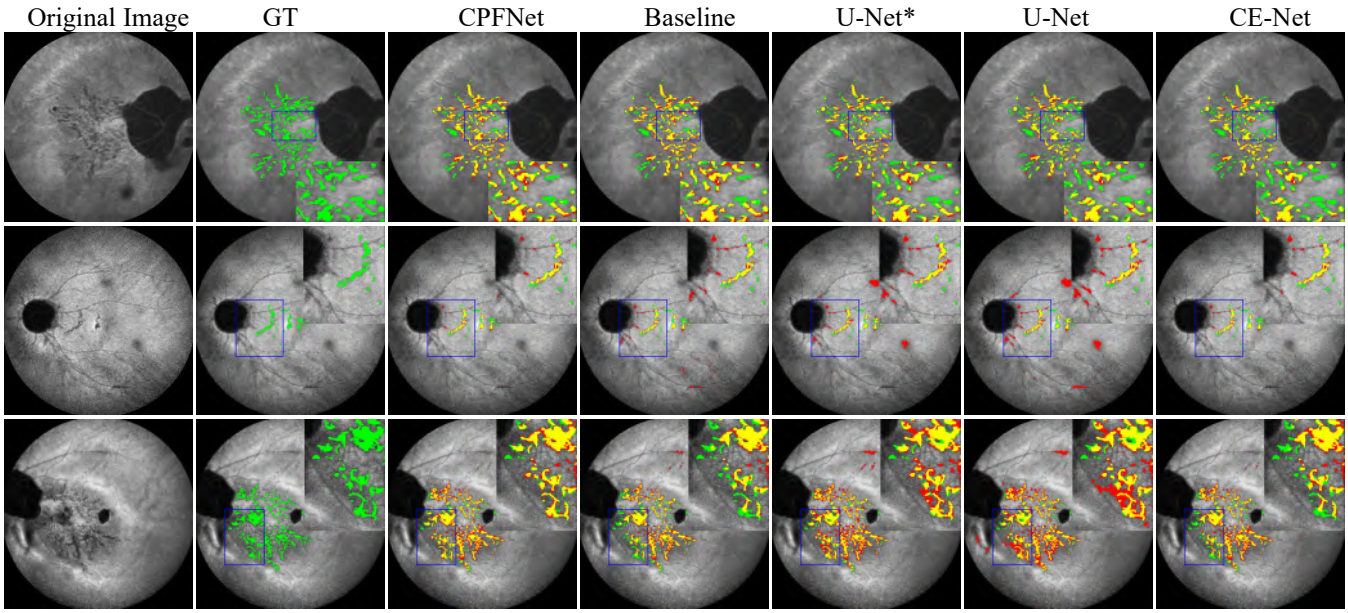

Fig.8. Examples of retinal linear lesion segmentation. The green, yellow and red regions represent the false negative (FN), true positive (TP) and false positive (FP) regions, respectively. From left to right: original image, ground truth (GT), our CPFNet, Baseline, U-Net* (represents U-Net+GPG+SAPF), U-Net and CE-Net.

TriamisuNet [41], cGAN [42], Attention U-Net [21], FastFCN [17], GCN [20], CE-Net [26] and our Baseline. As can be seen from Table II, FastFCN achieves a worst result with 25.62% for IoU. We think the possible reason is that some linear lesions are too small and FastFCN does not have reasonable skip connections to improve the resolution of the prediction map. CE-Net performs comparably with the proposed CPFNet in this task, while the performance of GCN is bad. Note that our CPFNet outperforms other methods and reaches 47.75%, 63.08%, 99.08%, and 71.77% in Jaccard index, Dice coefficient, Accuracy, Sensitivity and Specificity, respectively. Compared with the Baseline, the Jaccard index, Dice coefficient and Sensitivity remarkably increase by 3.50%, 3.71%, and 3.24%, respectively. Fig.8 shows some examples of linear lesion segmentation results with different approaches, which show our CPFNet is effective in this task and our proposed GPG and SAPF modules are robust.

C. Segmentation of Thoracic Organs at Risk

1) Overview

Radiation therapy is a selective treatment for cancer. In the radiotherapy procedure, the first step is to delineate the target tumor and the adjacent healthy organs, which are called as Organs at Risk (OAR) in CT images. Conventionally, the delineation is done manually by doctors, which is tedious and subjective. The automatic segmentation of OAR such as esophagus, heart, arteries, and trachea is especially challenging because the shape and position of OAR vary greatly between patients and the contours of OAR have low contrast in CT images. Several methods based on Generalized Hough Transform (GHT), atlas-registration [43] or level set [44] have been proposed for OAR segmentation. Recently, the OAR segmentation performance has been improved based on deep learning [4].

2) Dataset

We applied the proposed method on the CT dataset of Thoracic OAR from a public challenge: ISBI 2019 SegTHOR[3]. The thoracic OAR in this dataset include heart, aorta, trachea and esophagus. For different patients, the number of CT scan slices varies from 150 to 284 with a z-resolution from 2mm to 3.7mm. Each slice has 512×512 pixels with in-plane resolution varying from 0.90 mm to 1.37 mm per pixel. The most frequent resolution is $0.98\times0.98\times2.5$ mm^3. In this public challenge, 60 patients (11084 slices) are randomly split into a training set (40 patients, 7390 slices) and a testing set (20 patients, 3694 slices).

3) Data Processing and Experimental Setup

In order to reduce the irrelevant information and enhance contrast, pixel intensity normalization was performed, in which the intensity values of all scans were truncated to [-310,400] and linearly mapped to [0, 1]. In order to make use of the 3D contextual information, we transformed the 3D CT data into 2.5D to train our network. Specifically, three adjacent slices were stacked to form a 3-channel input data, and the network output the prediction of the middle slice. We also applied multiple data augmentations, including left-right flipping, up-down flipping, rotations of -15 degrees to 15 degrees and contrast normalization.

In addition, we randomly divided 40 training scans into 32 for training and 8 for validation. In order to make full use of the dataset, we added the validation data into the network for training in the last five epochs.

4) Results

In order to verify the reliability of our method, we submit the test results to the official challenge website for evaluation (evaluated with global Dice and Hausdorff distance). The results are listed in Table III. Without any post-processing, our method achieves remarkable results compared to the results of other submissions. Compared with Zhang [45], our method performs better for esophagus and heart segmentation, while a little worse for trachea and aorta segmentation, probably because we do not do any post-processing to clean up discontinuous false predictions between slices.

In addition, we also compare with some state-of-the-art CNNs, such as FCN [11], U-Net [7], and CE-Net [26]. As can be seen from Table III, our method still achieves excellent results. Although the second-ranked CE-Net [26] achieves comparable results with the proposed CPFNet, our method still has some improvement in average Dice and Hausdorff distance. Fig.9 shows some thoracic OAR segmentation results of different methods, which also demonstrate that the proposed CPFNet is suitable for OAR segmentation.

D. Retinal Edema Lesions Segmentation

1) Overview

In retinal optical coherence tomography (OCT) images, the segmentation of lesions such as retina edema area (REA), sub-retinal fluid (SRF) and pigment epithelial detachment (PED), is a crucial task for automated diagnosis of diabetic retinopathy. However, there are many challenges in multi-class lesion segmentation: 1) The boundary of the target is blurred and there is severe speckle noise in OCT images. 2) The data imbalance problem between different lesion categories is very severe. Previous studies [5], [6] were focused on single lesion segmentation of retinal edema, and the joint segmentation for these three lesions is still a blank.

2) Dataset

We acquired the dataset from a public competition: AI-challenger 2018 for automated segmentation of retinal edema lesions[4]. The dataset contains 85 retinal OCT cubes ($1024\times512\times128$) with ground truth. Due to the annotation problem, 83 OCT cubes with complete annotations were used in the experiments, which were divided into training set (40 cubes) and test set (43 cubes) according to the subjects. According to our statistics, PED lesion only accounts for 0.03% of the total area, which causes a serious class imbalance problem and makes the joint segmentation very hard. In order to improve the efficiency of network training, the OCT cube was resized to $512\times256\times128$. The 2.5D data processing and data augmentation method, the same as those applied in the segmentation of thoracic organs at risk, were also used in this experiment.

3) Result

We use the same indices as in the challenge to evaluate our approach, including Dice coefficient (Dice), Accuracy (Acc), Sensitivity (Sen) and Specificity (Spe). The proposed CPFNet is compared with eight other excellent networks such as FCN[11], U-Net[7], Attention U-Net [21], FastFCN[17] , MultiResUNet [2], DFN[19], GCN [20] and CE-Net[26]. As can be seen from Table IV, the proposed CPFNet achieves the best performance. The average Dice of the proposed CPFNet is 4.43% higher than MultiResUNet [21] without global context information, and the Dice coefficient for SRF segmentation is notably improved by 6.59% and reaches 83.49%. It is worth noting that although the CE-Net achieves comparable results with the CPFNet in REA and SRF segmentation, its PED segmentation performance is quite poor. We think that the reason may be that the pooling operation at the top of the CE-Net makes small PED regions indiscernible, and the same thing happens with DFN. The average Dice coefficient of

[3]https://segthor.grand-challenge.org

[4]https://challenger.ai/

TABLE III
PERFORMANCE COMPARISON OF THE SEGMENTATION FOR THORACIC ORGANS AT RISK

	Dice (%)					Hausdorff (mm)				
Methods	Ave	Esophagus	Heart	Trachea	Aorta	Ave	Esophagus	Heart	Trachea	Aorta
FCN [11]	0.7884	0.6003	0.9187	0.7736	0.8609	0.7690	1.3443	0.3080	0.9915	0.4323
U-Net [7]	0.8567	0.7693	0.9110	0.8536	0.8928	1.0681	1.0496	0.7356	1.5789	0.9084
Han [46]	0.8662	0.7518	0.9328	0.8884	0.8919	0.7269	0.9267	0.2184	0.6325	1.1300
Feng [47]	0.8746	0.7603	0.9401	0.8821	0.9159	0.3757	0.6862	0.1895	0.3647	0.2623
Zhang [45]	0.8835	0.7732	0.9384	**0.8939**	0.9285	0.5930	1.6774	0.2089	**0.2741**	0.2114
Mikhail [48]	0.8837	0.7986	0.9265	0.8850	0.9245	0.5313	0.6196	0.3002	0.9340	0.2712
CE-Net [26]	0.8848	0.7927	0.9448	0.8667	**0.9348**	0.4536	0.6994	0.1538	0.7871	**0.1741**
CPFNet	**0.8943**	**0.8120**	**0.9466**	0.8905	0.9282	**0.3562**	**0.4481**	**0.1460**	0.5423	0.2882

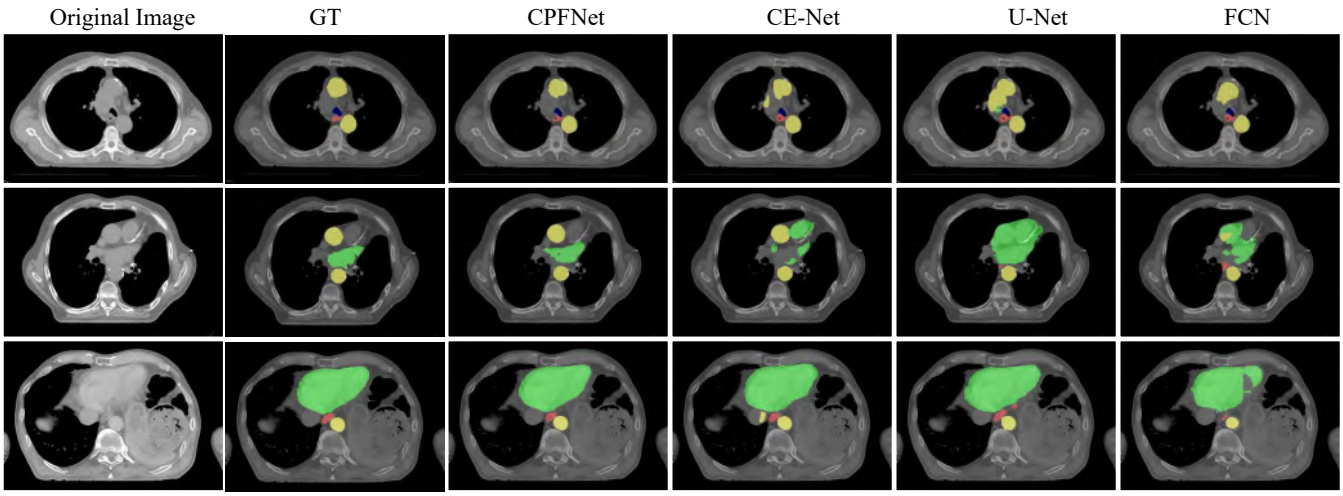

Fig.9. The segmentation examples of thoracic organs at risk. The blue, green, yellow and red regions represent the trachea, heart, aorta, esophagus, respectively. From left to right: original image (cropped for illustration), ground truth(GT. Because of no access of GT from the challenge website, for visual comparison, we manually annotated these three images carefully according to the annotations in the training set), our CPFNet, CE-Net, U-Net and FCN.

TABLE IV
PERFORMANCE COMPARISON OF RETINAL EDEMA LESION SEGMENTATION TASK

	Dice(%)				Sensitivity (%)			Specificity (%)			Accuracy (%)
Methods	Ave	REA	SRF	PED	REA	SRF	PED	REA	SRF	PED	Glob
DFN[19]	53.78	77.07	84.25	0.01	79.60	82.42	0.00	99.01	99.94	**100.00**	98.29
CE-Net [26]	54.57	80.93	82.78	0.00	86.01	80.84	0.00	99.03	99.94	**100.00**	98.24
FCN [11]	62.68	79.03	74.6	34.42	86.76	75.07	25.51	98.51	99.95	**100.00**	97.84
UNet [7]	69.91	75.3	77.27	57.15	87.92	78.50	61.57	97.96	99.92	99.98	97.38
Attention U-Net [21]	71.49	75.54	73.95	64.97	81.81	74.15	59.78	98.44	99.91	**100.00**	97.48
GCN [20]	73.28	74.99	76.27	68.60	83.21	78.56	67.61	98.41	99.91	**100.00**	97.46
FastFCN[17]	73.53	78.67	78.05	63.86	83.02	72.70	54.26	98.80	**99.96**	**100.00**	98.06
MultiResUNet [2]	75.42	76.53	76.90	72.85	81.27	77.98	77.20	98.98	99.91	**100.00**	97.76
Anatomynet [18]	77.57	**81.55**	77.37	73.79	88.43	72.42	77.83	98.68	**99.96**	99.99	98.25
Baseline	73.42	78.54	81.29	60.44	86.84	79.53	65.82	98.51	99.94	99.99	98.08
CPFNet	79.85	81.34	83.49	74.72	86.82	**83.13**	70.05	**99.18**	99.94	**100.00**	**98.34**
Baseline-50	76.90	79.55	77.37	73.79	88.43	72.42	77.83	98.68	**99.96**	99.99	98.25
CPFNet-50	**80.37**	80.74	**84.71**	**75.66**	**91.19**	82.78	**90.39**	98.38	99.94	99.99	97.98

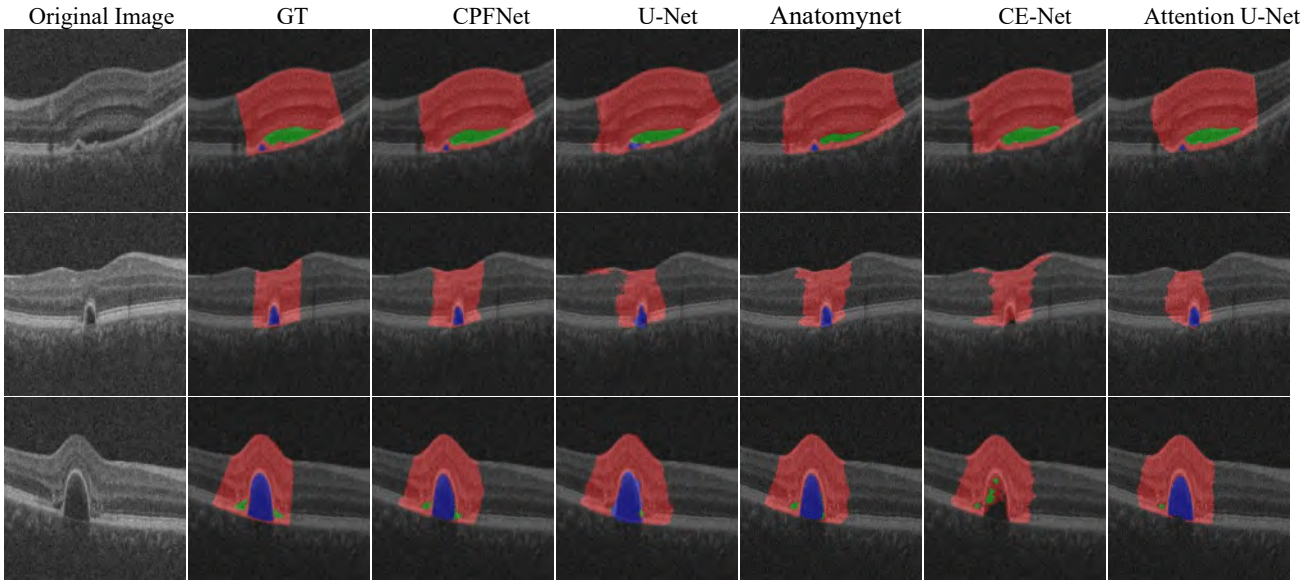

Fig.10. Examples of retinal edema lesion segmentation results with different methods, where the red, green and blue regions denote the REA, SRF and PED, respectively. From left to right: original image (cropped for illustration), ground truth (GT), our CPFNet, U-Net, Anatomynet, CE-Net and Attention U-Net.

Anatomynet is also worse than our method, although it achieves best Dice coefficient in REA. We think this is due to the fact that 3D CNN can improve continuity between slices.

In order to prove that our proposed modules can still perform well on network with more powerful backbone, we replace the ResNet34 with deeper backbone ResNet50 as encoder in the model. In Table IV, Baseline-50 represents Baseline with ResNet50 as backbone and CPFNet-50 represents CPFNet with ResNet50 as backbone. As can be seen from Table IV, although the improvement of our proposed network is reduced, there is still a considerable gap. The compelling improvement over the Baseline indicates that the insertions of two proposed GPG and SAPF modules are effective in context information capture and integration. The corresponding visual comparisons are shown in Fig.10.

IV. CONCLUSION

In this paper, we have proposed a novel deep learning framework CPFNet for medical image segmentation, which focuses on solving the weakness of global/multi-scale context information capture and integration in U-shape networks.

The proposed network adopts ResNet34 as the feature extractor, and two novel pyramidal modules including global pyramid guidance (GPG) module and scale-aware pyramid fusion (SAPF) module are designed and inserted into the U-Shape framework to exploit and fuse rich global/multi-scale context information.

In this paper, comprehensive experiments are performed on different types of medical image segmentation tasks to verify the effectiveness and generality of the proposed CPFNet, including skin lesion segmentation, retinal linear lesion segmentation, segmentation of thoracic organs at risk and retinal edema lesions segmentation. Although GCN has achieved comparable performance with the proposed CPFNet in skin lesion segmentation, it is unable to perform well broadly in other tasks such as retinal linear lesion segmentation and retinal edema lesions segmentation, which shows its lack of generality. Similarly, CE-Net performs comparably to the proposed CPFNet in the retinal linear lesion segmentation and the thoracic organ at risk segmentation challenge according to the most important Dice metric. However, CE-Net does not perform well on the skin lesion segmentation and retinal edema lesion segmentation challenge. Our proposed CPFNet has achieved good and consistent performances in these four different segmentation tasks, which suggests that the proposed CPFNet is more practicable and generalizable than other existing methods. Especially, our proposed GPG and SAPF modules are effective and universal, which can be easily introduced into other encoder-decoder network. We believe that our method can achieve better performance with further post-processing and can be extended to other medical image segmentation tasks, which is our near future work.

REFERENCES

[1] M. M. K. Sarker, H. A. Rashwan, F. Akram, S. F. Banu, A. Saleh, V. K. Singh *et al.*, "Slsdeep: Skin lesion segmentation based on dilated residual and pyramid pooling networks," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2018, pp. 21-29: Springer.

[2] N. Ibtehaz and M. S. Rahman, "MultiResUNet: Rethinking the U-Net Architecture for Multimodal Biomedical Image Segmentation," *arXiv preprint arXiv:1902.04049,* 2019.

[3] H. Jiang, X. Chen, F. Shi, Y. Ma, D. Xiang, L. Ye *et al.*, "Improved cGAN based linear lesion segmentation in high myopia ICGA images," *Biomedical optics express,* vol. 10, no. 5, pp. 2355-2366, 2019.

[4] R. Trullo, C. Petitjean, S. Ruan, B. Dubray, D. Nie, and D. Shen, "Segmentation of organs at risk in thoracic CT images using a sharpmask architecture and conditional random fields," in *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*, 2017, pp. 1003-1006: IEEE.

[5] S. J. Chiu, M. J. Allingham, P. S. Mettu, S. W. Cousins, J. A. Izatt, and S. Farsiu, "Kernel regression based segmentation of optical coherence tomography images with diabetic macular edema," *Biomedical optics express,* vol. 6, no. 4, pp. 1172-1194, 2015.

[6] F. G. Venhuizen, B. van Ginneken, B. Liefers, F. van Asten, V. Schreur, S. Fauser *et al.*, "Deep learning approach for the detection and quantification of intraretinal cystoid fluid in multivendor optical coherence tomography," *Biomedical optics express,* vol. 9, no. 4, pp. 1545-1569, 2018.

[7] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*, 2015, pp. 234-241: Springer.

[8] K. Zhou, Z. Gu, W. Liu, W. Luo, J. Cheng, S. Gao *et al.*, "Multi-Cell Multi-Task Convolutional Neural Networks for Diabetic Retinopathy Grading," in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2018, pp. 2724-2727: IEEE.

[9] D. Ciresan, A. Giusti, L. M. Gambardella, and J. Schmidhuber, "Deep neural networks segment neuronal membranes in electron microscopy images," in *Advances in neural information processing systems*, 2012, pp. 2843-2851.

[10] A. Sevastopolsky, "Optic disc and cup segmentation methods for glaucoma detection with modification of U-Net convolutional neural network," *Pattern Recognition and Image Analysis,* vol. 27, no. 3, pp. 618-624, 2017.

[11] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 3431-3440.

[12] V. Badrinarayanan, A. Kendall, and R. Cipolla, "Segnet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE transactions on pattern analysis and machine intelligence,* vol. 39, no. 12, pp. 2481-2495, 2017.

[13] T. Pohlen, A. Hermans, M. Mathias, and B. Leibe, "Full-resolution residual networks for semantic segmentation in street scenes," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 4151-4160.

[14] A. G. Roy, S. Conjeti, S. P. K. Karri, D. Sheet, A. Katouzian, C. Wachinger *et al.*, "ReLayNet: retinal layer and fluid segmentation of macular optical coherence tomography using fully convolutional networks," *Biomedical optics express,* vol. 8, no. 8, pp. 3627-3642, 2017.

[15] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," in *2016 Fourth International Conference on 3D Vision (3DV)*, 2016, pp. 565-571: IEEE.

[16] Y. Rong, D. Xiang, W. Zhu, F. Shi, E. Gao, Z. Fan *et al.*, "Deriving external forces via convolutional neural networks for biomedical image segmentation," *Biomedical optics express,* vol. 10, no. 8, pp. 3800-3814, 2019.

[17] H. Wu, J. Zhang, K. Huang, K. Liang, and Y. Yu, "FastFCN: Rethinking Dilated Convolution in the Backbone for Semantic Segmentation," *arXiv preprint arXiv:1903.11816,* 2019.

[18] W. Zhu, Y. Huang, L. Zeng, X. Chen, Y. Liu, Z. Qian *et al.*, "AnatomyNet: Deep learning for fast and fully automated whole-volume segmentation of head and neck anatomy," *Medical physics,* vol. 46, no. 2, pp. 576-589, 2019.

[19] C. Yu, J. Wang, C. Peng, C. Gao, G. Yu, and N. Sang, "Learning a discriminative feature network for semantic segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 1857-1866.

[20] C. Peng, X. Zhang, G. Yu, G. Luo, and J. Sun, "Large Kernel Matters--Improve Semantic Segmentation by Global Convolutional Network," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4353-4361.

[21] O. Oktay, J. Schlemper, L. L. Folgoc, M. Lee, M. Heinrich, K. Misawa *et al.*, "Attention U-Net: learning where to look for the pancreas," *arXiv preprint arXiv:1804.03999,* 2018.

[22] Y. Qin, K. Kamnitsas, S. Ancha, J. Nanavati, G. Cottrell, A. Criminisi *et al.*, "Autofocus layer for semantic segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2018, pp. 603-611: Springer.

[23] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 2881-2890.

[24] J.-J. Liu, Q. Hou, M.-M. Cheng, J. Feng, and J. Jiang, "A Simple Pooling-Based Design for Real-Time Salient Object Detection," *arXiv preprint arXiv:1904.09569,* 2019.

[25] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking atrous convolution for semantic image segmentation," *arXiv preprint arXiv:1706.05587,* 2017.

[26] Z. Gu, J. Cheng, H. Fu, K. Zhou, H. Hao, Y. Zhao *et al.*, "CE-Net: Context Encoder Network for 2D Medical Image Segmentation," *IEEE transactions on medical imaging,* 2019.

[27] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 7132-7141.

[28] L.-C. Chen, Y. Yang, J. Wang, W. Xu, and A. L. Yuille, "Attention to scale: Scale-aware semantic image segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 3640-3649.

[29] W. Sun and T. Wu, "Learning Spatial Pyramid Attentive Pooling in Image Synthesis and Image-to-Image Translation," *arXiv preprint arXiv:1901.06322,* 2019.

[30] X. Li, W. Wang, X. Hu, and J. Yang, "Selective Kernel Networks," *arXiv preprint arXiv:1903.06586,* 2019.

[31] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770-778.

[32] H. Li, P. Xiong, J. An, and L. Wang, "Pyramid attention network for semantic segmentation," *arXiv preprint arXiv:1805.10180,* 2018.

[33] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1251-1258.

[34] Available: https://distill.pub/2016/deconv-checkerboard/

[35] A. Kardynal and M. Olszewska, "Modern non-invasive diagnostic techniques in the detection of early cutaneous melanoma," *Journal of dermatological case reports,* vol. 8, no. 1, p. 1, 2014.

[36] M. E. Celebi, Q. Wen, H. Iyatomi, K. Shimizu, H. Zhou, and G. Schaefer, "A state-of-the-art survey on lesion border detection in dermoscopy images," *Dermoscopy Image Analysis,* pp. 97-129, 2015.

[37] N. C. Codella, D. Gutman, M. E. Celebi, B. Helba, M. A. Marchetti, S. W. Dusza *et al.*, "Skin lesion analysis toward melanoma detection: A challenge at the 2017 international symposium on biomedical imaging (isbi), hosted by the international skin imaging collaboration (isic)," in *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, 2018, pp. 168-172: IEEE.

[38] P. Tschandl, C. Rosendahl, and H. Kittler, "The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions," *Scientific data,* vol. 5, p. 180161, 2018.

[39] S. Bass and J. Sherman, "The Retinal Atlas," *Optometry - Journal of the American Optometric Association,* vol. 82, no. 3, pp. 136-137, 2011.

[40] K.-C. Hung, M.-S. Chen, C.-M. Yang, S.-W. Wang, and T.-C. Ho, "Multimodal imaging of linear lesions in the fundus of pathologic myopic eyes with macular lesions," *Graefe's Archive for Clinical and Experimental Ophthalmology,* vol. 256, no. 1, pp. 71-81, 2018.

[41] S. Jégou, M. Drozdzal, D. Vazquez, A. Romero, and Y. Bengio, "The one hundred layers tiramisu: Fully convolutional densenets for semantic segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2017, pp. 11-19.

[42] M. Mirza and S. Osindero, "Conditional generative adversarial nets," *arXiv preprint arXiv:1411.1784,* 2014.

[43] M. Han, J. Ma, Y. Li, M. Li, Y. Song, and Q. Li, "Segmentation of organs at risk in CT volumes of head, thorax, abdomen, and pelvis," in *Medical Imaging 2015: Image Processing*, 2015, vol. 9413, p. 94133J: International Society for Optics and Photonics.

[44] E. Schreibmann, D. M. Marcus, and T. Fox, "Multiatlas segmentation of thoracic and abdominal anatomy with level set-based local search," *Journal of applied clinical medical physics,* vol. 15, no. 4, pp. 22-38, 2014.

[45] L. Zhang, L. Wang, Y. Huang, and H. Chen, "Segmentation of thoracic organs at risk in ct images combining coarse and fine network," *http://ceur-ws.org/Vol-2349/SegTHOR2019_paper_5.pdf,* 2019.

[46] S. Kim, Y. Jang, K. Han, H. Shim, and H. J. Chang, "A Cascaded Two-step Approach For Segmentation of Thoracic Organs," *http://ceur-ws.org/Vol-2349/SegTHOR2019_paper_3.pdf,* 2019.

[47] M. Feng, W. Huang, Y. Wang, and X. Yuxia, "Multi-organ segmentation using simplified dense v-net with post processing," *http://ceur-ws.org/Vol-2349/SegTHOR2019_paper_11.pdf,* 2019.

[48] K. Vladimir, D. Dmitry, P. Artem, and B. Mikhail, "Segmentation of thoracic organs at risk in ct images using localization and organ-specific cnn," *http://ceur-ws.org/Vol-2349/SegTHOR2019_paper_9.pdf,* 2019.